

Prediction of cervical spondylosis classification using deep learning with convolutional neural network

Tachi H^{1,2}, Kokabu T¹, Sudo H², Iwasaki N²

¹ Department of Orthopaedic Surgery, Eniwa Hospital, Eniwa, Hokkaido, Japan, ² Department of Orthopaedic Surgery, Hokkaido University Graduate School of Medicine, Sapporo, Hokkaido, Japan

Email of Presenting Author: hitachi198885@gmail.com

INTRODUCTION: Cervical spondylotic myelopathy (CSM) is becoming more common in the initial consultation to non-specialists because of the increasing prevalence of the disease with aging society. It is recommended that CSM patients with progressive symptoms consult with specialists and undergo early surgery. However, it is difficult for non-specialist to determine whether surgery is necessary. It would be a significant advantage for non-specialists if CSM with surgery and cervical spondylotic radiculopathy (CSR) without surgery can be classified from initial X-ray using deep learning algorithm (DLA). The purpose of this study was to validate prediction accuracy of the DLA for classifying into CSM for surgery and CSR for conservative therapy from the cervical spine X-ray.

METHODS: We enrolled 150 patients who underwent surgery for CSM and 150 patients who underwent conservative therapy for CSR between January 2016 and November 2022. A lateral radiograph of each patient in the neutral position was obtained, which was then labeled as either CSM or CSR. The radiographs were cropped to C3-6 vertebrae as region of interest images (Figure 1). The preprocessed image data were inputted to the convolutional neural network (CNN) as training datasets and internal validation datasets. Six-fold cross validation was performed for training and internal validation to avoid selection bias. Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize the areas of diagnostic evidence for the CNN model. To evaluate the performance of the trained DLAs using an independent external validation dataset, other 100 images including 50 CSM images and 50 CSR images were collected. The images were classified into CSM and CSR by the trained network with weight and bias of the six trained DLAs.

RESULTS: Table 1 showed the performance parameters of the CNN model in the internal validation, respectively. Total dataset had the accuracy of 0.86 and AUC of 0.92. For the external validation using average prediction of the 6 trained DLAs, AUC of the ROC curve was 0.99, where the prediction accuracy, sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio and negative likelihood ratio were 0.96, 0.98, 0.94, 0.94, 0.98, 16.33, and 0.02, respectively. The Grad-CAM (heat map) showed that DLAs mainly extracted high feature regions such as the anterior vertebral osteophytes, degenerative change in vertebral endplates and narrowing of the spinal canal (Figure 2).

DISCUSSION: The DLAs had high accuracy with adequate capture of high feature regions classifying CSM with surgery and CSR without surgery from cervical spine X-ray. This suggests that an application utilizing the DLAs would be helpful for non-specialists in consulting with specialists even if the clinic had no MRI. To complete the application with the DLAs using uncropped images, a large number of images for CSM and CSR need to be collected.

SIGNIFICANCE/CLINICAL RELEVANCE: The present study confirmed prediction accuracy of the DLA for classifying into CSM for surgery and CSR for conservative therapy from the cervical spine X-ray.

ACKNOWLEDGEMENTS: This work was supported by the Japan Agency for Medical Research and Development (JP18he1302026h0003).

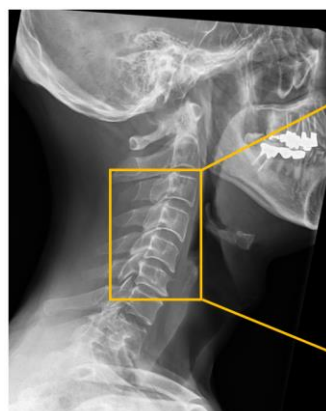


Figure 1. Pre-processing of input images.

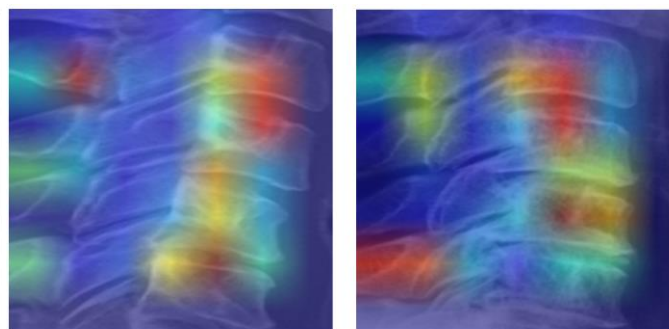


Figure 2. Grad-CAM in CNN model.

Table 1. Parameters in six-fold cross validation.

	Sensitivity	Specificity	PPV	NPV	Accuracy	PLR	NLR
Dataset 1	0.88	0.88	0.88	0.88	0.88	7.58	0.14
Dataset 2	0.80	0.87	0.80	0.87	0.84	6.00	0.23
Dataset 3	0.85	0.81	0.91	0.72	0.84	4.55	0.18
Dataset 4	0.87	0.74	0.74	0.87	0.8	3.35	0.18
Dataset 5	0.87	0.85	0.83	0.88	0.86	5.87	0.15
Dataset 6	0.92	0.92	0.92	0.92	0.92	11.08	0.08
Total	0.87	0.85	0.85	0.86	0.84	5.65	0.16

PPV= positive predictive value, NPV= negative predictive value, PLR= positive likelihood ratio, NLR= negative likelihood ratio. NA means not applicable.